Introduction to Web Technologies

Journalism 303: stuff you’ll use from the web

Martin Frigaard

2021-09-21

Data Journalism Web Technologies

Overview

1.) What is data journalism?

1.1) Definitions

1.2) Examples

2.) Code files for data journalism

2.1) R Code

2.2) HTML

2.3) CSS

3) Data file formats

3.1) .CSV

3.2) .XML

3.3) .JSON

What is data journalism?

1.1) Wikipedia:

> "*a journalistic process based on analyzing and filtering large data sets for the purpose of creating or elevating a news story*"

1.2) fivethirtyeight:

> "*Data-driven news and analysis*"

1.3) the upshot (NYT):

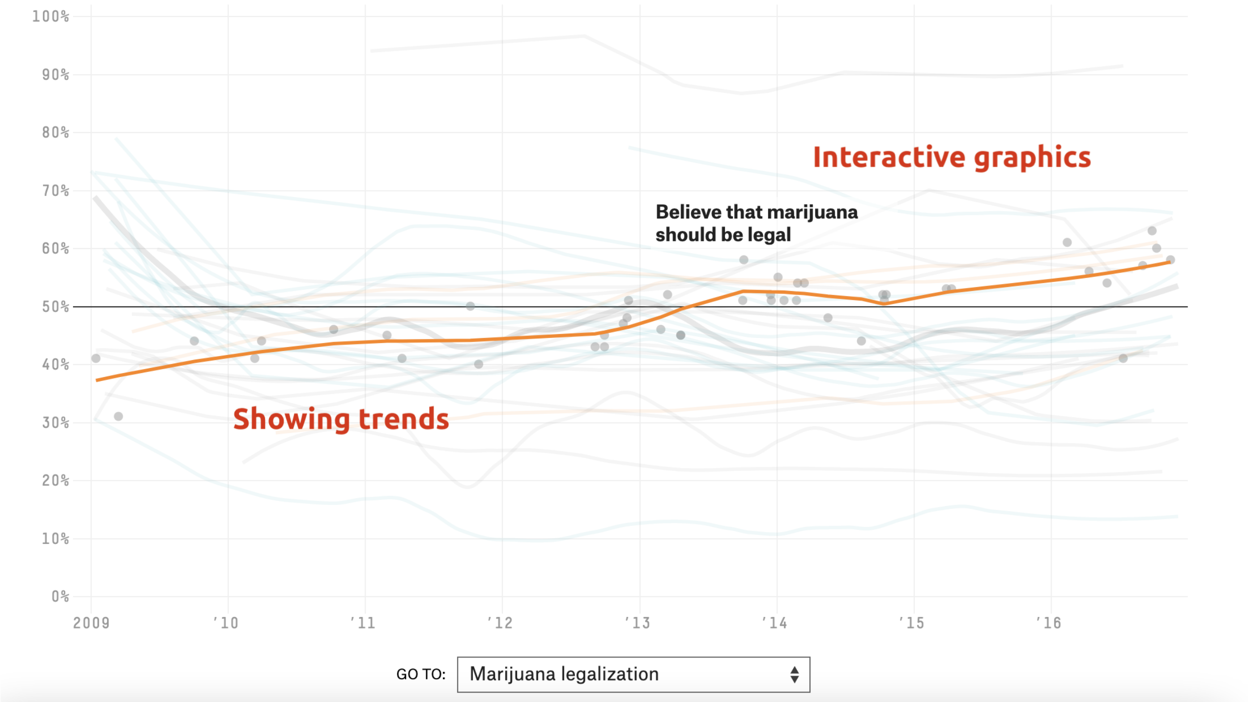
> "*Analytical journalism in words and graphics* "

What is data journalism? (Example 1)

FiveThirtyEight's *How America’s Thinking Changed Under Obama: Public opinion on 32 big issues over the past eight years*

What is data journalism? (Example 1)

FiveThirtyEight's *How America's Thinking Changed Under Obama: Public opinion on 32 big issues over the past eight years*

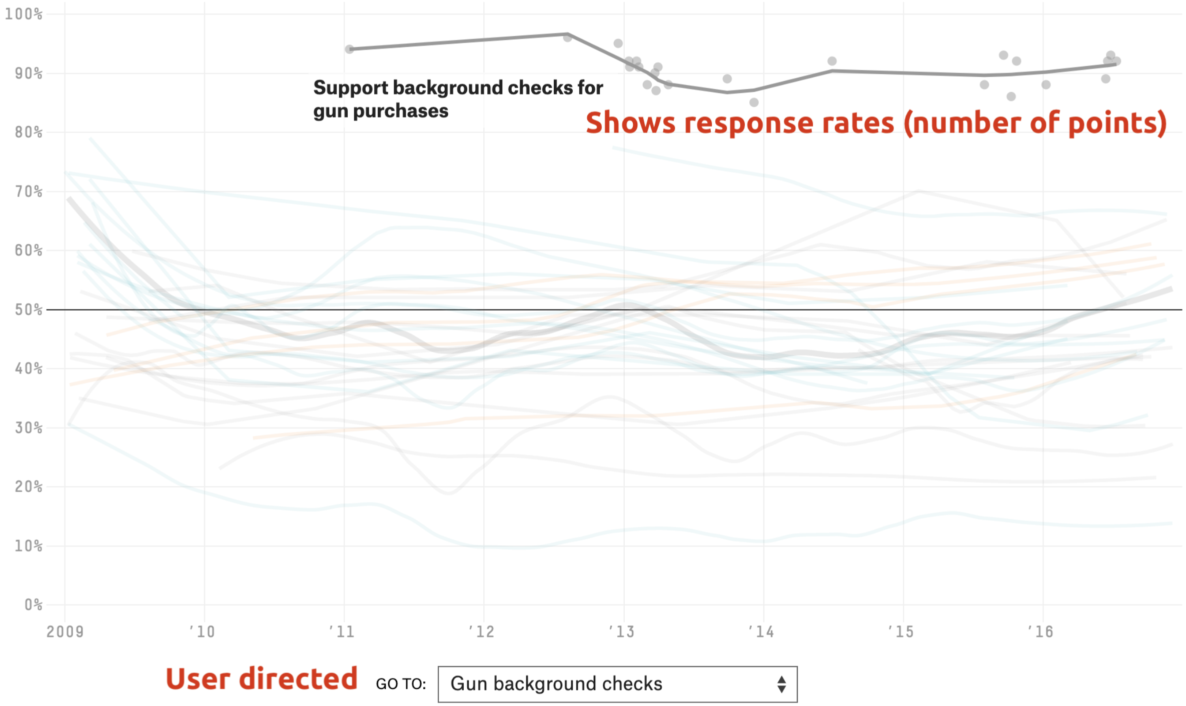


Interactive graphics

Showing trends

What is data journalism? (Example 1)

FiveThirtyEight's *How America’s Thinking Changed Under Obama: Public opinion on 32 big issues over the past eight years*



Shows response rates (number of points)

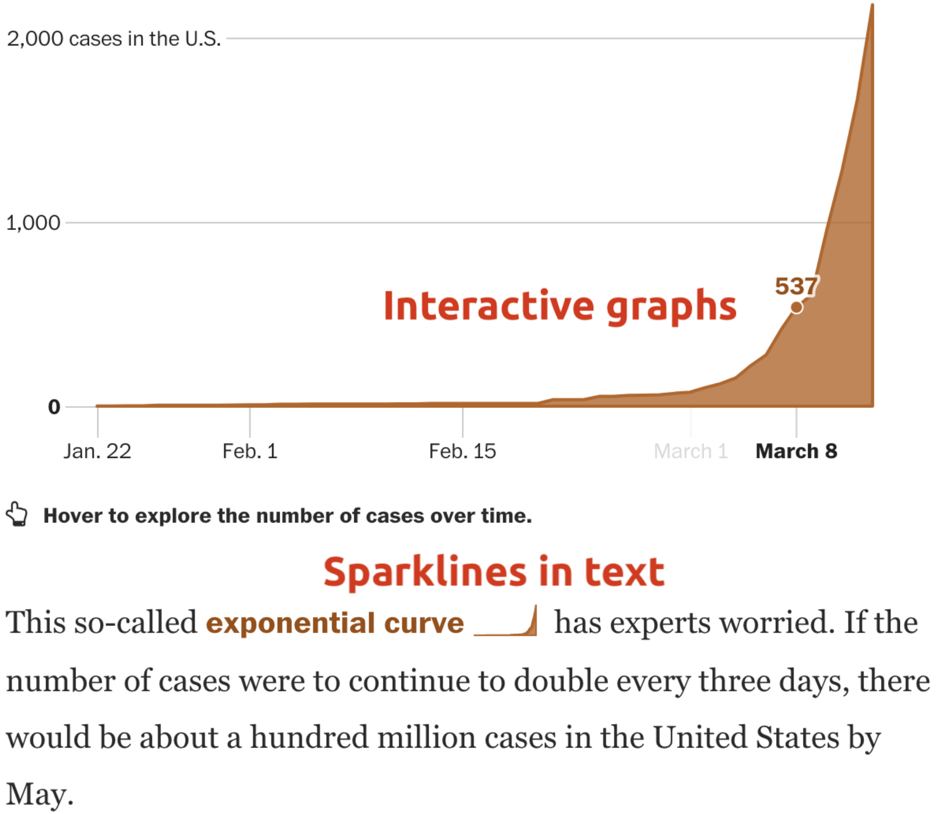
User directed

What is data journalism? (Example 2)

Washington Post's *Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve"*

What is data journalism? (Example 2)

> Washington Post's *Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve"*



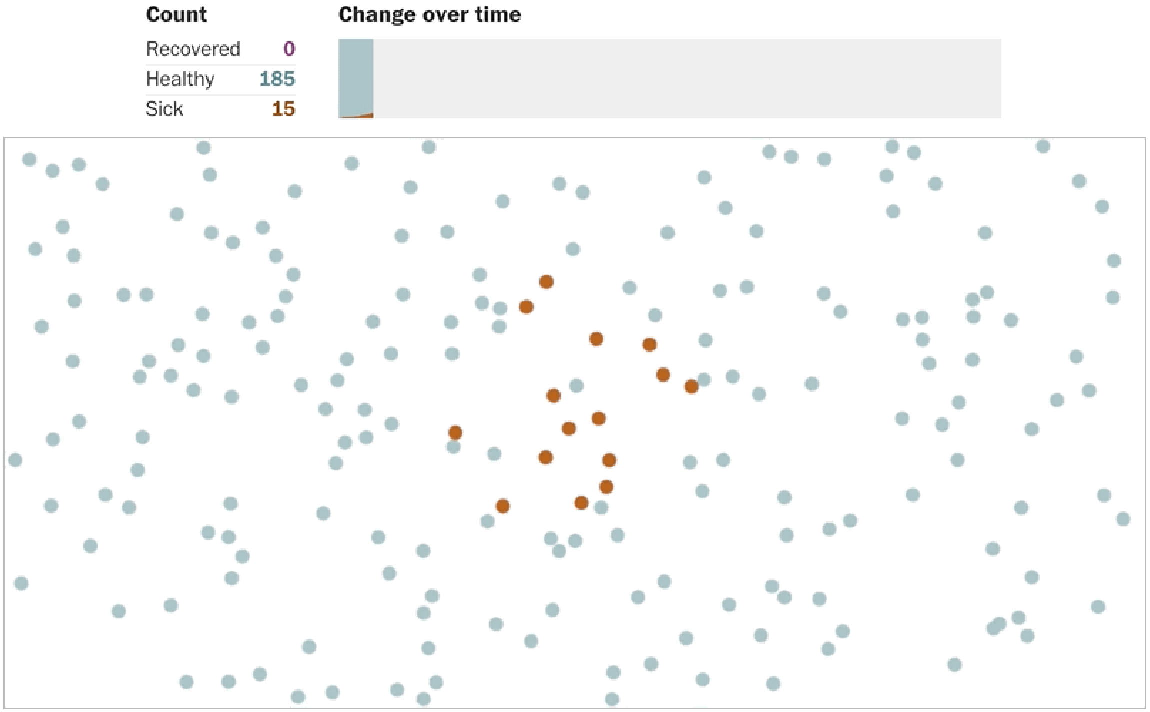
Hover to explore the number of cases over time.

Sparklines in text

This so-called **exponential curve** has experts worried. If the number of cases were to continue to double every three days, there would be about a hundred million cases in the United States by May.

What is data journalism? (Example 2)

> Washington Post's *Why outbreaks like coronavirus spread exponentially, and how to "flatten the curve"*



Count

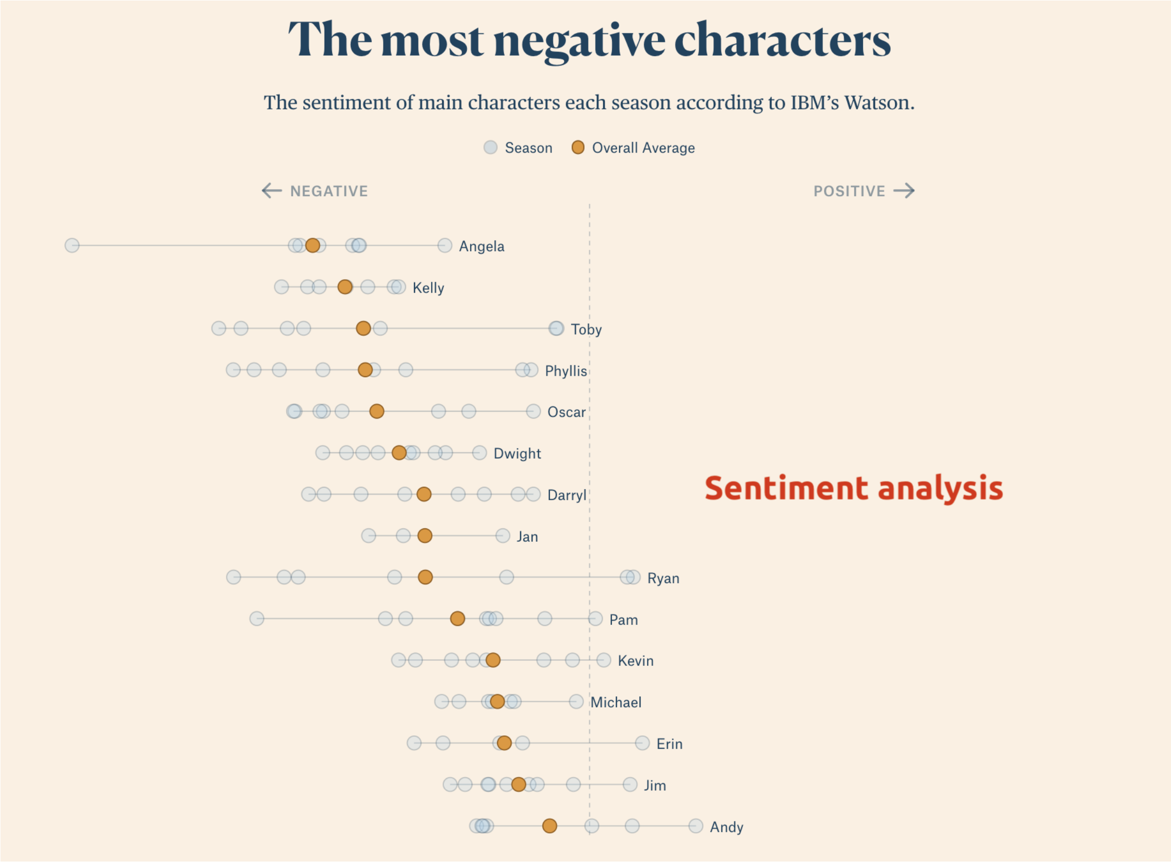
Recovered	0
Healthy	185
Sick	15

What is data journalism? (Example 3)

The Pudding's *‘The Office’ Dialogue in Five Charts: A breakdown of how every character contributed to the show.*

What is data journalism? (Example 3)

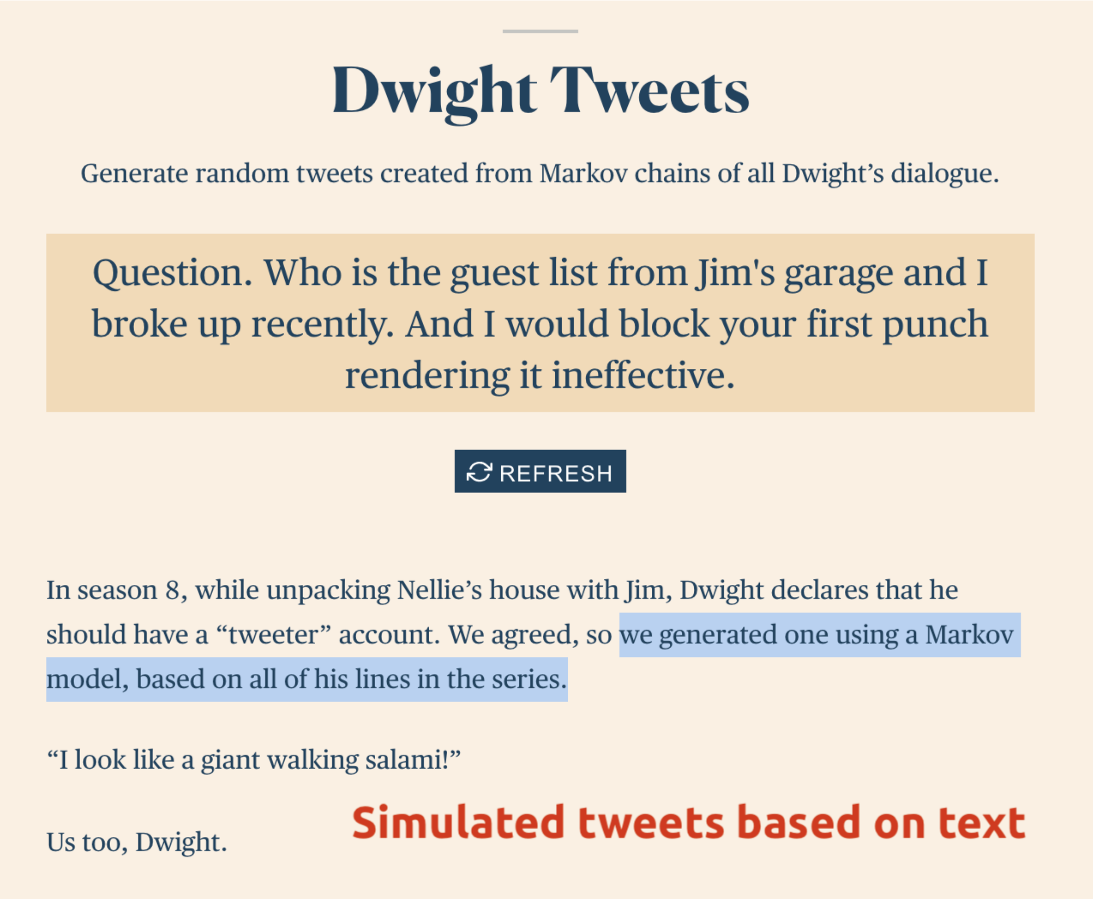
> The Pudding's *‘The Office’ Dialogue in Five Charts: A breakdown of how every character contributed to the show.*



Sentiment analysis

What is data journalism? (Example 3)

> The Pudding's *'The Office' Dialogue in Five Charts: A breakdown of how every character contributed to the show.*



Dwight Tweets

Generate random tweets created from Markov chains of all Dwight's dialogue.

Question. Who is the guest list from Jim's garage and I broke up recently. And I would block your first punch rendering it ineffective.

In season 8, while unpacking Nellie's house with Jim, Dwight declares that he should have a "tweeter" account. We agreed, so we generated one using a Markov model, based on all of his lines in the series.

"I look like a giant walking salami!"

Us too, Dwight.

Simulated tweets based on text

Code for Data Journalists

`Code` file formats and extensions

We can determine the language of code by it's **`file`** extension

- **`file.R`** = R code file (or script)

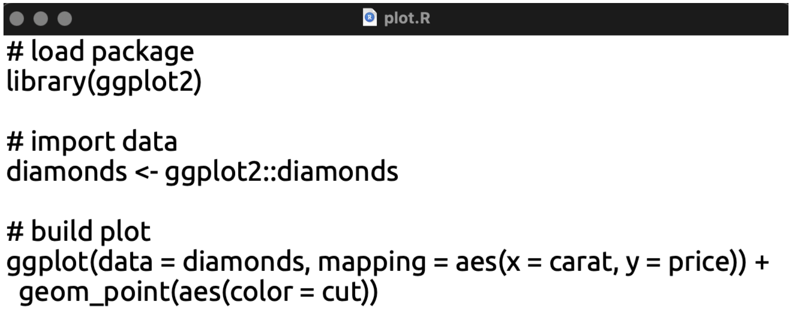
- **`file.html`** = HTML code file (or webpage)

- **`file.css`** = CSS code file (or stylesheet)

`Code` file formats and extensions

Text editors (like Notepad or TextEdit) can be used to view code files

Below is the **`plot.R`** code file in TextEdit



```
# load package
library(ggplot2)

# import data
diamonds <- ggplot2::diamonds

# build plot
ggplot(data = diamonds, mapping = aes(x = carat, y = price)) +
  geom_point(aes(color = cut))
```

R Code

Why do we need to write code?

1) R is an actual *language*, which means it gives us the ability to express our ideas with precision.

2) R code is text, so we can use copy + paste and Google

R Code: grammar & syntax

- A code's *syntax* defines the rules for it's grammar and punctuation

- The characters and words have specific meanings (just like in English)

R Code: grammar & syntax

- Characters and words have to be written in a particular order for R code work

- In the R, there are two primary components to the grammar (or sytnax): *functions* and *objects*

R `Code`: R functions & objects

`functions` are like *verbs*

`verb(noun)`

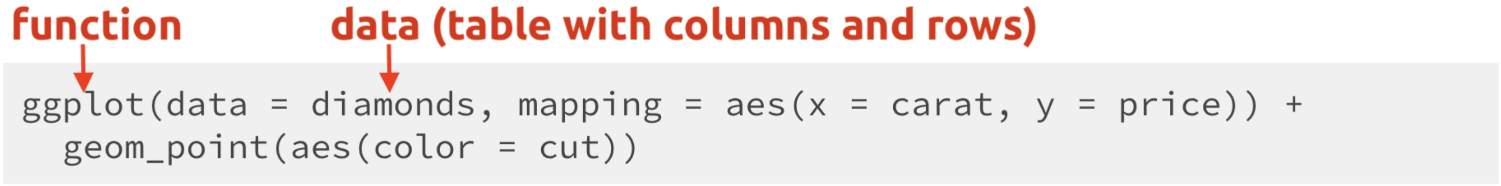
is like...

`objects` are like *nouns*

`function(object)`

R Code: A quick example

Here is some example R code for building a graph:

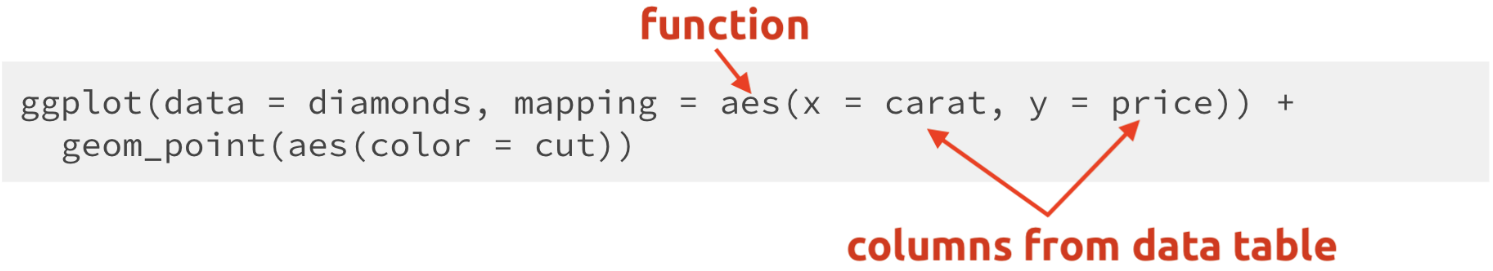


```
ggplot(data = diamonds, mapping = aes(x = carat, y = price)) +
  geom_point(aes(color = cut))
```

> The functions (`ggplot()`) *'do things'* to the objects (`diamonds`)

R Code: A quick example

Here is some example R code for building a graph:

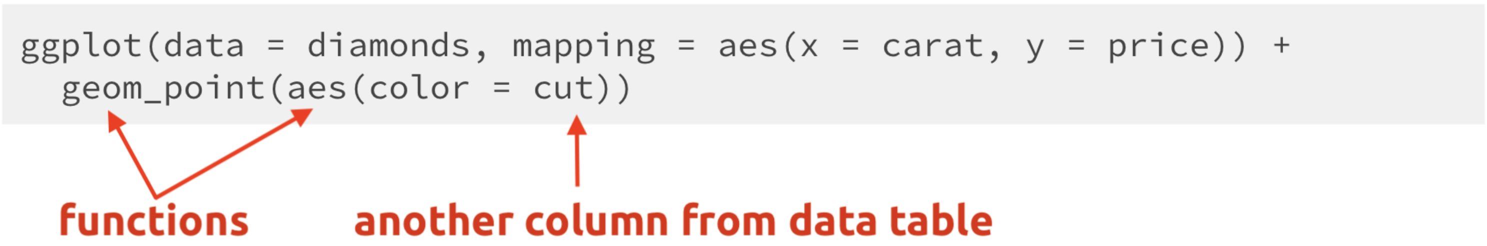


```
ggplot(data = diamonds, mapping = aes(x = carat, y = price)) +
  geom_point(aes(color = cut))
```

> *We're telling R we want to use the `diamonds` data, and we want the `carat` column on the `x` axis, and the `price` column on the `y`.*

R Code: A quick example

Here is some example R code for building a graph:



```
ggplot(data = diamonds, mapping = aes(x = carat, y = price)) +
  geom_point(aes(color = cut))
```

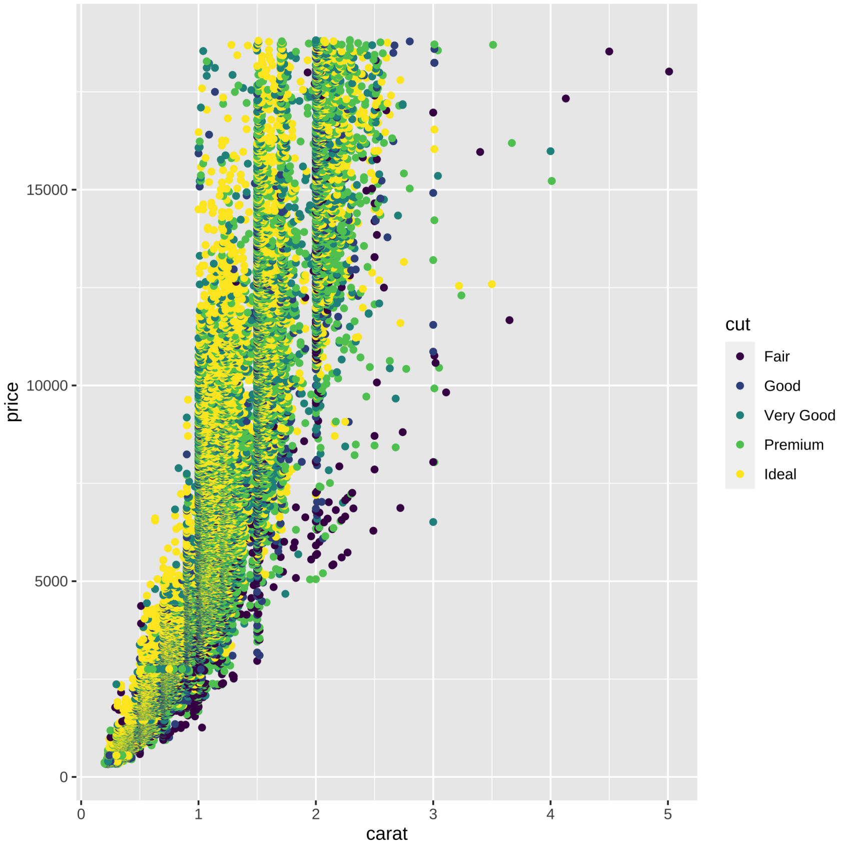
functions

another column from data table

> *We want the graph to have 'points' (or dots), so we use the `geom_point()` function, and we want the points colored by the `cut` column.*

R Code: A quick example

```
ggplot(data = diamonds, mapping = aes(x = carat, y = price)) +
  geom_point(aes(color = cut))
```



Code for Data Journalists

HTML

HTML

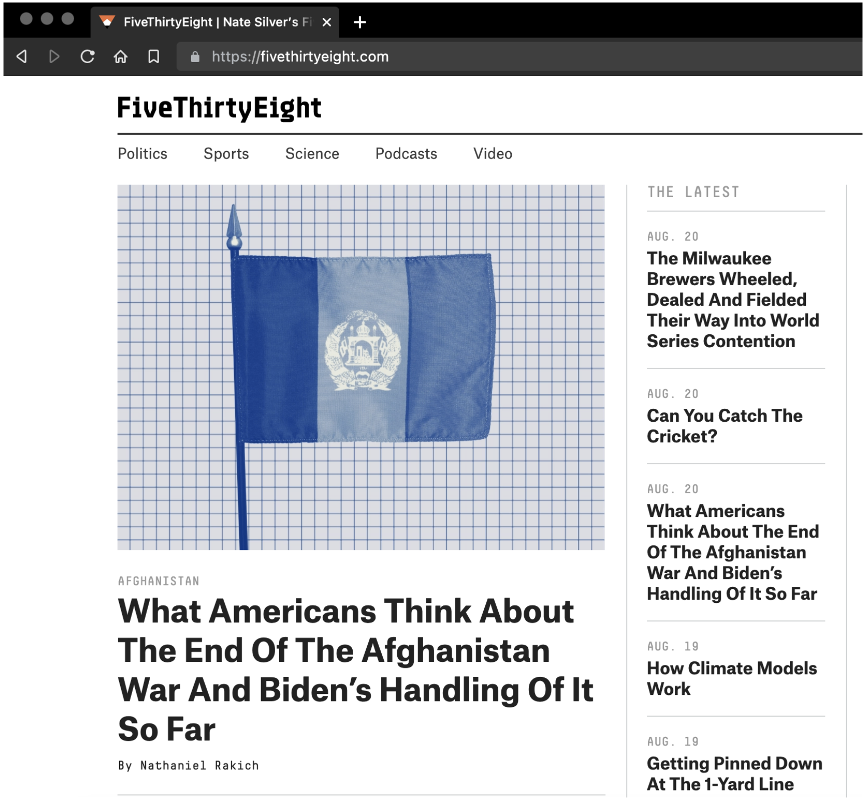
HTML stands for 'HyperText Markup Language' and is a computer language used to create web pages

HTML code can be run by opening the file containing the code with any web browser (Chrome, Safari, Firefox, etc.)

HTML5 is the current standard

HTML

Head over to the fivethirtyeight landing page


FiveThirtyEight | Nate Silver's Fi
https://fivethirtyeight.com

FiveThirtyEight

Politics Sports Science Podcasts Video

AFGHANISTAN

What Americans Think About The End Of The Afghanistan War And Biden's Handling Of It So Far

By Nathaniel Rakich

THE LATEST

AUG. 20
The Milwaukee Brewers Wheeled, Dealed And Fielded Their Way Into World Series Contention

AUG. 20
Can You Catch The Cricket?

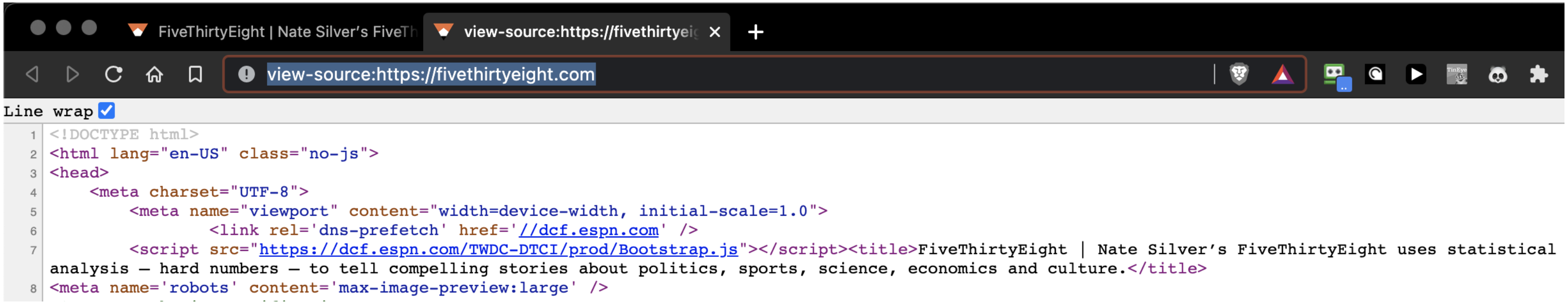
AUG. 20
What Americans Think About The End Of The Afghanistan War And Biden's Handling Of It So Far

AUG. 19
How Climate Models Work

AUG. 19
Getting Pinned Down At The 1-Yard Line

HTML

Right click on the page and click 'view source'


FiveThirtyEight | Nate Silver's FiveTh
view-source:https://fivethirtyei
view-source:https://fivethirtyeight.com

Line wrap ☑

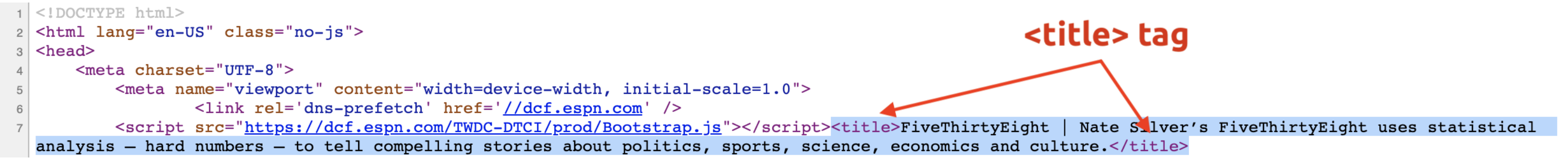
```
<!DOCTYPE html>
<html lang="en-US" class="no-js">
<head>
    <meta charset="UTF-8">
        <meta name="viewport" content="width=device-width, initial-scale=1.0">
                <link rel='dns-prefetch' href='//dcf.espn.com' />
        <script src="https://dcf.espn.com/TWDC-DTCI/prod/Bootstrap.js"></script><title>FiveThirtyEight | Nate Silver’s FiveThirtyEight uses statistical analysis — hard numbers — to tell compelling stories about politics, sports, science, economics and culture.</title>
<meta name='robots' content='max-image-preview:large' />
```

HTML: structure

HTML consists of **elements** and **tags**

Elements have a start tag, followed by the element content, followed by an end tag

```
<element>content</element>
```



```
<!DOCTYPE html>
<html lang="en-US" class="no-js">
<head>
    <meta charset="UTF-8">
        <meta name="viewport" content="width=device-width, initial-scale=1.0">
                <link rel='dns-prefetch' href='//dcf.espn.com' />
        <script src="https://dcf.espn.com/TWDC-DTCI/prod/Bootstrap.js"></script><title>FiveThirtyEight | Nate Silver’s FiveThirtyEight uses statistical
analysis — hard numbers — to tell compelling stories about politics, sports, science, economics and culture.</title>
```

HTML: `elements` and `tags`

Start tag:

```
<title>
```

Content:

```
<title>FiveThirtyEight | Nate Silver’s FiveThirtyEight uses statisti...
```

End tag:

```
<title>FiveThirtyEight | Nate Silver’s FiveThirtyEight uses...</title>
```

HTML: attributes

Attributes appear in the start tag and are of the form **`attribute="attribute value"`**

The code below shows the start tag for an **`img`** element, with an attribute called **`src`** with a value **`"image.png"`**

```
<img src="image.png">
```

HTML: tags to know

```
<!-- html comments (not read or displayed by browser) -->
<!DOCTYPE html> <!-- document type declaration -->
<html lang="en-US"> <!-- describes the web page -->
  <head> <!-- header of the HTML document -->
    <title></title> <!-- title of the HTML document -->
  </head>
  <body> <!-- visible page content -->
    <h1> <!-- level 1 header (others include h2-h6) -->
    <!-- href is url, followed by displayed text -->
      <a
      href="https://www.website.com">A link to website
      </a>
      <!-- src the path to an image file -->
        <img src="example.com/example.jpg">
    <h1/>
  </body>
</html>
```

Most code is added in the `<body>` *element*

Code for Data Journalists

css

CSS

CSS stands for 'Cascading Style Sheets'

CSS is used for describing the layout, colors, and fonts of a HTML document

CSS: structure

CSS syntax

```
<style>
    h1 {
        color: blue;
    }
</style>
```

- `<style>` start tag
  - a **selector** (`h1`)
  - an open bracket (`{`)
    - **property name** (`color`)
    - colon (`:`)
    - **property value** (`blue`)
    - semi-colon (`;`)
  - a closed bracket (`}`)
- `</style>` end tag

`CSS`: use

`CSS` is most useful when included in an external CSS file (i.e., `my_style_sheet.css`)

We can then reference the `my_style_sheet.css` style sheet using the `<link>` tag

```
<link href="my_style_sheet.css" rel="stylesheet"
      type="text/css">
```

Data file formats

`Data` file types

Data comes in a variety of formats, but this course will focus on 'plain text formats'

Text editors can read and write plain text files

Plain text files are portable across different computer operating systems

CSV: 'comma-separated values' files

The first line is a "header" and contains column (or variable) names in each of the fields (using letters, digits, and underscores)

Each following line represents a new row (or observation)

Any field *may be* quoted, but fields with embedded commas or double-quote characters *must be* quoted

CSV: 'comma-separated values' files

How .csv files look in text editors:

```
name, age, street, city, state, zip
Sally, 24, 6 Taylor Ave., Commack, NY, 11725
Fred, 38, 450 Grant Ave., Fort Dodge, IA, 50501
Deb, 48, 661 Spring Drive, Phillipsburg, NJ, 08865
```

How .csv files look in a spreadsheet:

name	age	street	city	state	zip
Sally	24	6 Taylor Ave.	Commack	NY	11725
Fred	38	450 Grant Ave.	Fort Dodge	IA	50501
Deb	48	661 Spring Drive	Phillipsburg	NJ	08865

`XML`: Extensible Markup Language

`XML` consists of of XML elements with a start tag and an end tag (with plain text content or other XML elements in-between)

The start tag may include attributes of the form `attribute="value"` (case-sensitive)

All attribute values must be enclosed within double-quotes

XML: structure

Below is a small XML document

```
<?xml version="2.0"?>
<heights>
<filename>heights.txt</filename>
<case date="24-JAN-2019"
      height="78.9"/>
</heights>
```

- **`root element`** = `<heights>`
- **`start tag`** = `<filename>`
- **`content`** = `heights.txt`
- **`end tag`** = `</filename>`
- **`element name`** = `case`
- **`attribute name`** = `date`
- **`attribute value`** = `"24-JAN-2019"`
- **`attribute name`** = `height`
- **`attribute value`** = `"78.9"`

The root element is the `heights` element with `filename` and `case` elements nested within the `heights`

`JSON`: JavaScript Object Notation

`JSON` is a lightweight data storage format similar in structure to `XML` but different syntax/format

Common format for data from application programming interfaces (APIs)

`JSON`: structure

Data are stored as:

- Numbers (double)

- Strings (double quoted)

- Boolean (true or false)

- Array (ordered, comma separated enclosed in square brackets`[]`)

- Object (unorderd, comma-separated collection of key:value pairs in curley brackets `{}`)

JSON: structure

Recall the .csv format:

```
name, age, street, city, state, zip
Sally, 24, 6 Taylor Ave., Commack, NY, 11725
Fred, 38, 450 Grant Ave., Fort Dodge, IA, 50501
Deb, 48, 661 Spring Drive, Phillipsburg, NJ, 08865
```

Same data as JSON:

```
[
  {
    "name": "Sally",
    "age": 24,
    "street": "6 Taylor Ave.",
    "city": "Commack",
    "state": "NY",
    "zip": "11725"
  },
  {
    "name": "Fred",
    "age": 38,
    "street": "450 Grant Ave.",
    "city": "Fort Dodge",
    "state": "IA",
    "zip": "50501"
  },
  {
    "name": "Deb",
    "age": 48,
    "street": "661 Spring Drive",
    "city": "Phillipsburg",
    "state": "NJ",
    "zip": "08865"
  }
]
```

Recap

Data journalists use programming languages as a tool to process, store, and display data

Code is the preferred technology because it's a language and allows us to be precise and expressive

Plain text data file formats are simple, lowest-common-denominator storage formats

Data in a plain text formats are usually arranged in rows, with several values on each row